AI-Assisted Expert Computable Operation Definition (CODef) Development for Real-World Research

Michael Buck,[1] Craig G. Parker,[1] Aaron W. C. Kamauu[1]

[1]Navidence Inc., Salt Lake City, Utah, USA

Why is this important?

Objective: As regulatory agencies look to require computable operational definitions (CODefs) as part of submissions from life sciences companies, it is essential to have ways to create and maintain this content quickly and accurately.

Problem: The process to develop code list/value sets that accurately represent clinical concepts for multiple EHR, claims and life science data sets can be very time consuming.

What was the process we followed?

We developed a novel three-step process to assist expert informaticists using Artificial Intelligence (AI).

Claude.ai → **Claude.ai** → **Perplexity.ai / Human Review**

1. Leveraging Claude.ai we processed and created a concept list from the National Comprehensive Cancer Network (NCCN) Gastric Cancer guidelines.

2. Leveraging Claude.ai we linked the concepts to common terminologies, like ICD-10-CM, LOINC, CPT, and RxNorm.

3. Assisted by Perplexity.ai a non-physician informaticist verified the accuracy of the concept and terminology mappings.
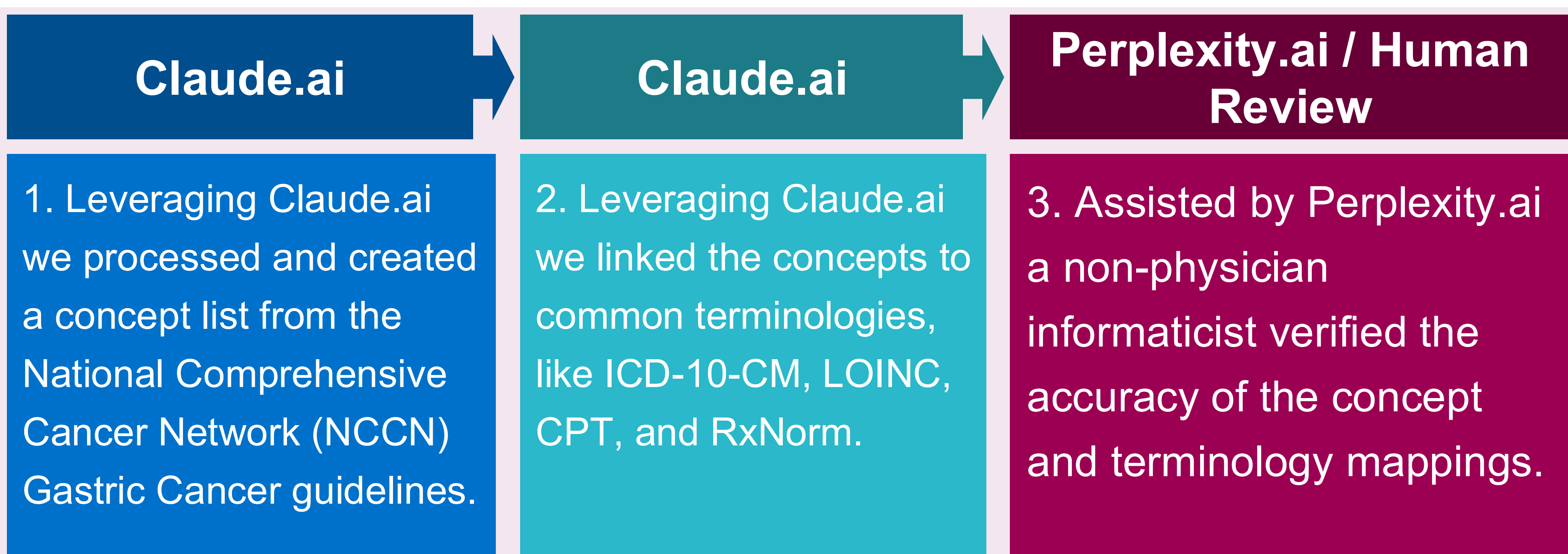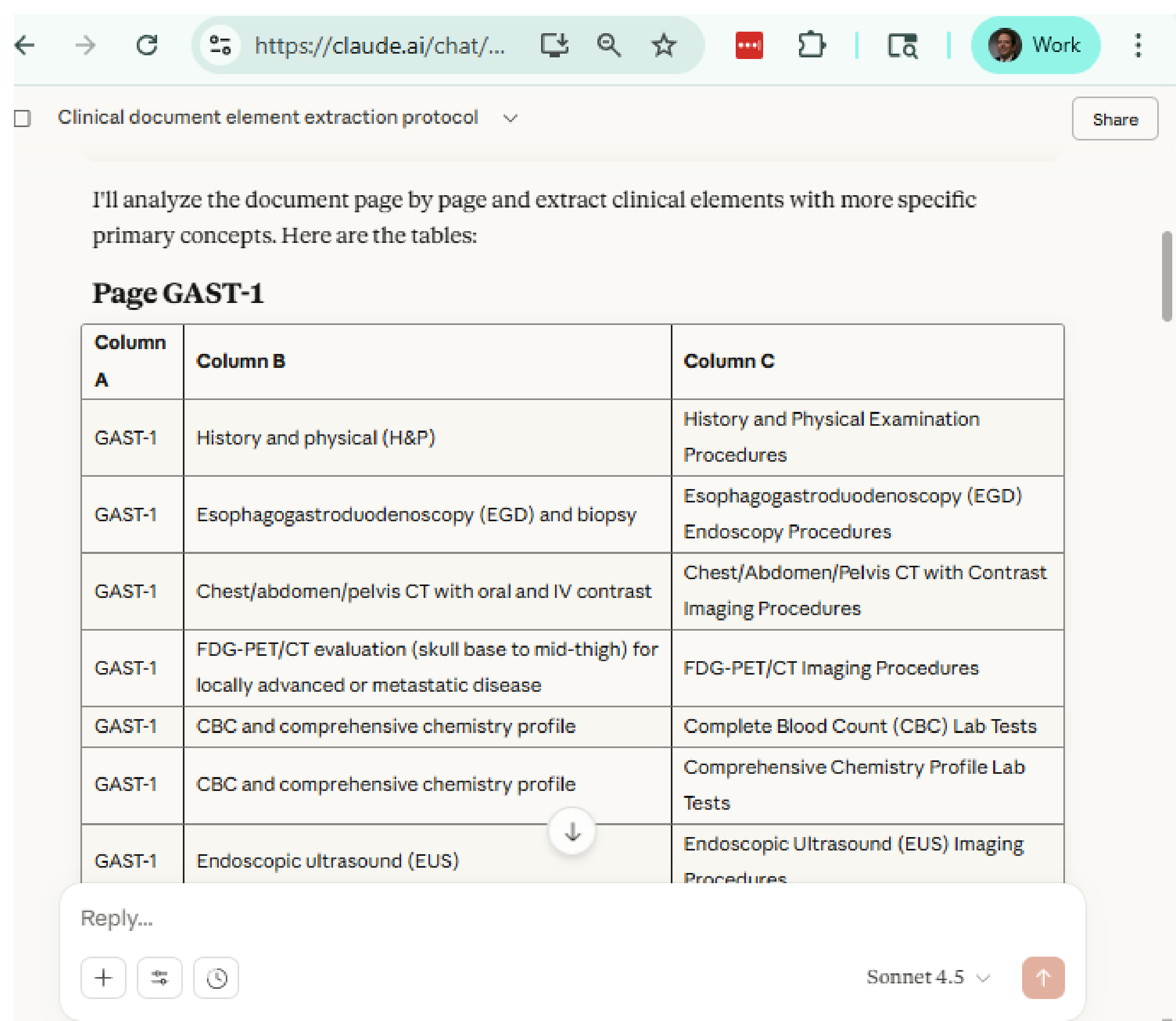
### Figure 1: Creation of Concept List leveraging Claude.ai

AI PROMPT: For each page, please produce a table for each page that analyzes the attached document and performs the following tasks: 1. Extract all clinical elements as exact text quotes. 2. Maintain the original numbering or order of the elements as they appear in the document. 3. For each extracted element, identify its primary concept. Present your results in a table format with page number as such: Column A: file page number [page number found in lower right corner like GAST-1] Column B: Source wording [Exact quote of the element] Column C: Proposed label [Build a label with the following structure: Primary Concept and Clinical Element Type. The Clinical Element types include Diagnoses, Procedures, Therapies (for medications), Lab Tests and Observations (e.g., Diabetes Diagnoses, Biopsy Procedures, Albumin Lab Tests, Systolic Blood Pressure Observations)] Repeat this format for each clinical element found in the document, preserving their original sequence. Do not paraphrase or summarize the element; only use the exact text as it appears in the document for the Source wording. Redo this analysis and update the Primary Concepts to be more specific using more specific details from the Source Wording like Esophagogastroduodenoscopy (EGD) Endoscopy Procedures and Fluorouracil and oxaliplatin Chemotherapy Therapies.

Figure 2: Concept linking to Common Terminologies like RxNorm

Figure 3: Human / Perplexity.ai Concept Review

Gastric Laboratory Test Concept	LOINC	LOINC Label	Informaticist Review
Blood Chemistry Laboratory Tests	1751-7	Albumin [Mass/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2160-0	Creatinine [Mass/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2345-7	Glucose [Mass/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	6768-6	Alkaline phosphatase [Enzymatic activity/volume] in Serum o	yes
Blood Chemistry Laboratory Tests	1920-8	Aspartate aminotransferase [Enzymatic activity/volume] in Se	yes
Blood Chemistry Laboratory Tests	1742-6	Alanine aminotransferase [Enzymatic activity/volume] in Seru	yes
Blood Chemistry Laboratory Tests	1975-2	Bilirubin.total [Mass/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2885-2	Protein [Mass/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2823-3	Potassium [Moles/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2951-2	Sodium [Moles/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2075-0	Chloride [Moles/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	2028-9	Carbon dioxide, total [Moles/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	3016-3	Thyrotropin [Units/volume] in Serum or Plasma	yes
Blood Chemistry Laboratory Tests	33747-0	Hepatitis B surface antigen [Presence] in Serum	Not real LOINC
Blood Chemistry Laboratory Tests	22314-9	Hepatitis A virus IgM Ab [Presence] in Serum	yes
Post-Treatment Blood Testing Laboratory Tests	1751-7	Albumin [Mass/volume] in Serum or Plasma	yes
Post-Treatment Blood Testing Laboratory Tests	2160-0	Creatinine [Mass/volume] in Serum or Plasma	yes
Post-Treatment Blood Testing Laboratory Tests	2345-7	Glucose [Mass/volume] in Serum or Plasma	yes
Post-Treatment Blood Testing Laboratory Tests	718-7	Hemoglobin [Mass/volume] in Blood	yes
Post-Treatment Blood Testing Laboratory Tests	4544-3	Hematocrit [Volume Fraction] of Blood by Automated count	yes
Post-Treatment Blood Testing Laboratory Tests	777-3	Platelets [#/volume] in Blood by Automated count	yes
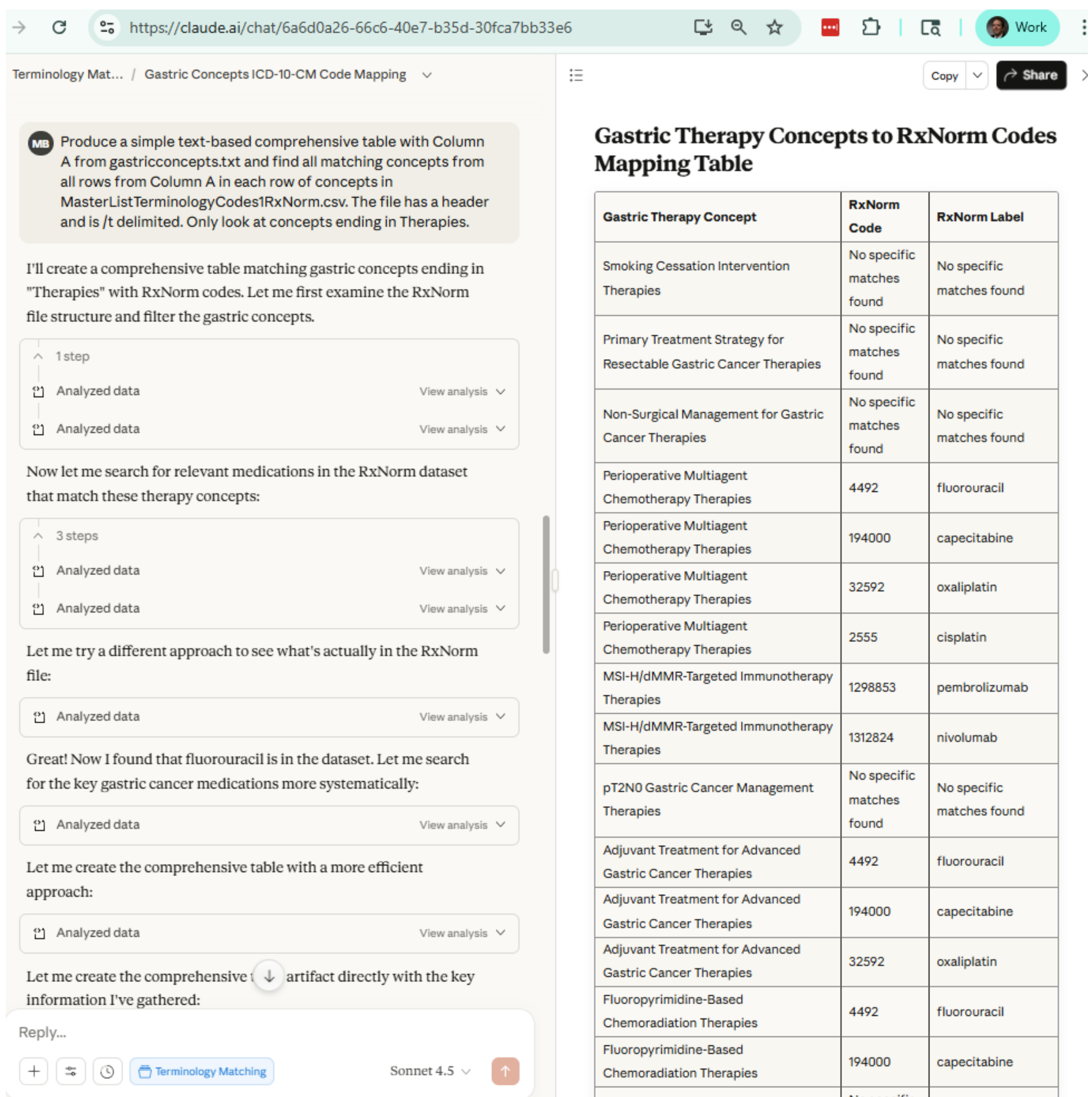
Results:

- CPT procedures: 101 distinct concepts mapped to 86 code terms with 54 concepts missing a mapped term & 24 concepts with incorrect terms for gastric cancer.
- ICD-10-CM diagnoses: 9 distinct concepts mapped to 11 terms with no concepts missing a mapping & all 9 verified as correct.
- RxNorm medications: 81 distinct concepts mapped to 33 terms with 9 concepts missing a mapped term & all terms with correct codes/text, except 12 with hallucinated RxNorm codes but correct medication text.
- LOINC labs: 3 distinct concepts mapped to 33 terms with 0 concepts missing a mapped term & all terms with correct codes/text except 1 with a hallucinated LOINC code but correct lab text.

Conclusions

Conclusions: AI was able to accelerate an expert human informaticist's ability to create and maintain value sets (for diagnoses, medications, and labs) for CODefs quickly and accurately and which are compliant with expectations from regulators / stakeholders.